Histogramme

Réflexion sur une représentation graphique particulière parfois abusivement utilisée tant dans l'enseignement que dans l'application de la statistique.

Jean-Claude Régnier

Nous avons choisi ici de développer notre point de vue pédagogique à propos de la notion d'histogramme. Ce qui nous y incite n'est à chercher ni dans un attachement affectif particulier à cette notion ni dans une admiration pour les propriétés géométriques et algébriques de cette représentation mais dans le fait que cette notion outil du domaine de la statistique est un objet d'enseignement. En devenant objet d'enseignement nous pouvons nous interroger sur les déformations subies par le concept du domaine de la statistique mathématique dans ce processus de transposition didactique[1]. Le terme histogramme est largement utilisé dans divers manuels scolaires, dans certains logiciels tels que des *tableurs* ou des logiciels de statistique.

Les questions principales qui nous guident sont :

Dans quels buts prévoit-on d'enseigner la notion d'histogramme dès la classe de quatrième ?

Quel(s) sens recouvre la notion d'histogramme pour l'enseignant quand elle est effectivement enseignée ? Quel(s) sens prend cette notion chez l'apprenant ?

En admettant que cette notion doive être enseignée, quelles caractéristiques minimales présentes dans le concept doivent être préservées dès l'initiation ? Quelles situations didactiques peut-on construire pour faire apprendre ce concept ?

Nous allons tenter d'exposer nos propres réponses dont nous postulons *a priori* la réfutabilité

Quelques caractéristiques du concept d'histogramme en statistique mathématique.

Tout d'abord quelle peut être l'étymologie du mot *histogramme* ? Le mot se compose du grec *histo* « tissu, texture, trame » et *gramme* « dessin, trace, écrit ». Selon le Grand Robert, le mot *histogramm* apparaît en

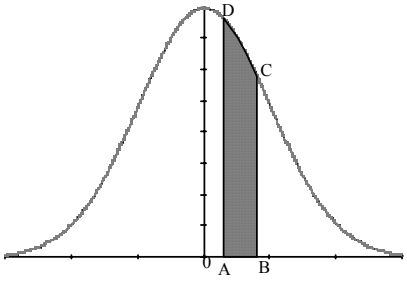
[1] *Transposition didactique, du savoir savant au savoir enseigné* Chevallard, Y., La pensée sauvage, RDM, 1991, 233p

anglais vers 1903. Notons que *hist-, histo-* et *histio*[2]- renvoient à une racine analogue or en grec *histion* désigne « voile de navire, tenture ou toile». Ce mot ne figure ni dans le *Dictionnaire de l'Académie Française* (1822) ni dans l'*Encyclopédie du XIXème Siècle* (1858) ni dans les volumes de mathématiques de l'*Encyclopédie Méthodique* de D'Alembert et Diderot (1789). En revanche notons que le terme *diagramme* très utilisé figure dans ces trois dictionnaires. D'Alembert & Diderot le définissent ainsi « *en géométrie, c'est une figure ou une construction de lignes, destinée à l'explication ou à la démonstration d'une proposition. Ce mot est plus d'usage en latin, diagramma, qu'en français; on se sert simplement du mot figure* ». Cette définition demeure actuelle. Nous y recourrons pour résoudre notre problème terminologique. L'emploi de l'*histogramme* en tant que représentation graphique serait attribué[3] à A. Guerry en 1833. Une interprétation possible serait qu'une forme fréquente d'histogramme évoque celle de la voile triangulaire d'un bateau. Évidemment cette conjecture devrait être confrontée à ce qu'en pensaient les premiers utilisateurs de ce mot ... dans leurs écrits!

Qu'est-ce que peut être un histogramme ?

Nous allons tenter d'en donner une définition mathématique la plus proche possible du concept sans pour autant développer tous les outils mathématiques qui se trouvent impliqués.

Considérons X une variable statistique (*resp.* une variable aléatoire) continue dont la loi de fréquence (*resp.* la loi de probabilité) est caractérisée par une fonction densité f. L'histogramme est alors la surface délimitée par la courbe représentative de f et l'axe des abscisses.



(figure 1)

exemple de fonction densité f(t) $= \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})$

dont la fonction de répartition est

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{t^2}{2})dt$$

[2] Histiodromie = art de la navigation par le moyen des voiles (Dictionnaire de l'Académie française 1822)

[3] *Histoire de la Statistique* , Droesbeke, J.J, & Tassi, Ph, PUF Que sais-je ? n°2527--1990 p 8

Ainsi la fréquence des observations appartenant à un intervalle [a ; b [est définie par :

$$P(\{ \omega \in \Omega ; X(\omega) \in [a ; b [\}) = P([a ; b[) = F(b)-F(a) = \int_a^b f(t)dt$$ où F est la fonction de répartition admettant la fonction densité f comme fonction dérivée.

De même nous avons $\int_{-\infty}^{+\infty} f(t)dt = 1$ ou encore

$$P(]-\infty ; b[) = \int_{-\infty}^{b} f(t)dt$$

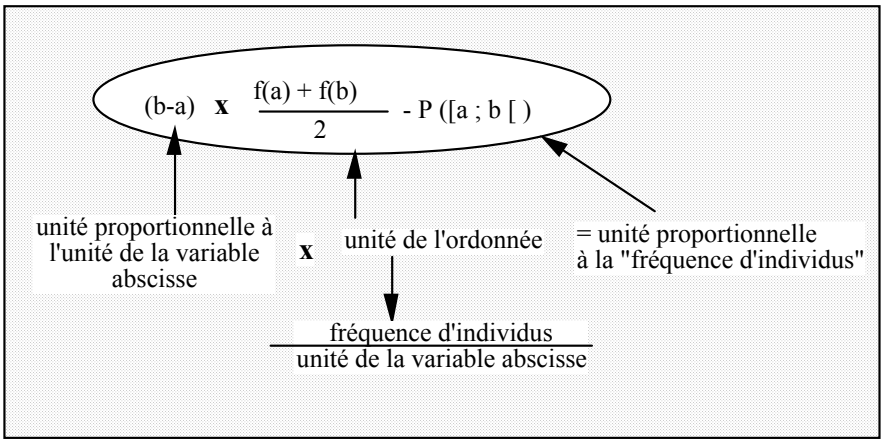
Ainsi dans l'histogramme c'est l'aire située sous la courbe qui est à prendre en compte. Cette aire s'obtient à l'aide du calcul intégral. Calculer une probabilité ou une fréquence , c'est donc mesurer une aire délimitée par la courbe de densité.

Intuitivement l'unité de la variable-abscisse est celle de la variable étudiée. Celle de l'aire est une "fréquence d'individus". Ainsi l'unité de la variable-ordonnée est alors "fréquence d'individus / unité de la variable". Observons le domaine ABCD dans la figure précédente. Approximativement[4] nous pouvons le considérer comme un trapèze rectangle rectiligne en estimant que l'arc de courbe CD est proche d'un segment rectiligne. Son aire peut être approchée par la formule suivante $A = (b-a)\frac{f(a)+f(b)}{2}$ qui pourrait être interprétée comme la fréquence des observations appartenant à l'intervalle [a ; b[. Remarquons encore que si nous rapprochons le point A du point B alors le point C se rapproche du point D et à la *limite* le trapèze se réduit à un segment dont la mesure de la surface est égale à 0. Ceci traduit géométriquement une caractéristique des variables statistiques continues qui fournit un résultat que l'entendement humain a sans doute des difficultés à appréhender par l'intuition.

Revenons à une démarche utilisée par le physicien avec les équations aux dimensions :

---

[4] Des précautions doivent être prises à l'égard de cette approximation dont la qualité est tributaire de la proximité des points A et B choisis ainsi que de la forme de la courbe reflétée par les variations de la fonction dérivée f'. Il ne s'agit ici que de suggérer une image grossière.



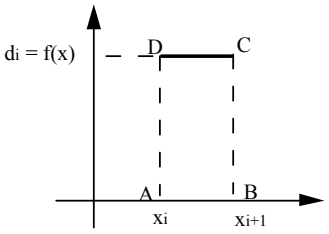
Ainsi la représentation graphique *histogramme* doit être placée dans un repère tel que

variable – abscisse	unité de la variable étudiée
variable - ordonnée	unité densité de fréquence ou de probabilité
variable -surface	unité : fréquence, probabilité ou effectif

Quand rencontre-t-on cette notion en statistique ?

Dans les études statistiques portant sur une variable quantitative continue telle par exemple la masse mesurée en kg de billes d'acier produites par une machine. Il s'agit d'une mesure au sens du physicien. Les données peuvent être placées raisonnablement dans des intervalles dont l'amplitude correspond à la précision de l'instrument de mesure. Prenons par exemple l'hectogramme. Après quoi, nous choisissons un nombre fini de ces intervalles que nous notons $[x_i ; x_{i+1}[$ avec i = 1 à k. Le choix du nombre d'intervalles est tout à fait discutable. Nous en percevons vite l'arbitraire et il faudrait en analyser les conséquences chaque fois. Mais *a priori* tout nombre positif réel dans un intervalle dont l'amplitude est déterminée par les contraintes physiques de l'objet fabriqué (par exemple : les billes ne peuvent dépasser la masse de 1,2 kg) peut convenir. Ajoutons qu'il pourrait être utile de prévoir une extension en considérant les deux intervalles particuliers, les deux demi-droites $[x_k ; + \infty [$ et $] - \infty; x_1[$. En ce sens pour modéliser le contrôle du réglage de la machine, il paraît pertinent de choisir le modèle d'une variable continue définie par une loi de fréquence absolument continue c'est à dire définie par une densité de fréquence.

Pour diverses raisons qui pourraient être discutées, nous pouvons considérer que sur un intervalle toutes les valeurs ont *la même chance d'être* le résultat d'une mesure. Ceci se traduit, dans le modèle mathématique adopté, *par* le fait que la densité de fréquence est constante sur un intervalle. Le graphique de la figure 2 traduit cette idée)



(figure 2)

Comment pourrait-on estimer les valeurs d_i ?

En effet ce que le modèle postule *a priori,* c'est que la loi de fréquence est uniforme sur chaque intervalle avec une densité définie comme suit :

$f(x) = d_i$ sur $[x_i ; x_{i+1}[$ avec i = 1 à k et $f(x) = 0$ sur $[x_{k+1} ; + \infty [$ et $] - \infty; x_1[$. On procède alors à un sondage en mesurant le plus grand nombre possible d'objets dans des conditions supposées identiques. La résultat de chaque mesure correspondant à un des k intervalles construits *a priori*. Les n mesures seront réparties entre les intervalles et nous obtiendrons :

intervalles	$] - \infty; x_1[$	$[x_1; x_2[$	...	$[x_i ; x_{i+1}[$	...	$[x_k ; x_{k+1}[$	$[x_{k+1}; + \infty [$
effectifs	$n_0 = 0$	n_1	...	n_i	...	n_k	$n_{k+1} = 0$
fréquences	$f_0 = 0$	f_1	...	f_i	...	f_k	$f_{k+1} = 0$

Rappelons que $\sum_{i=1}^{k} n_i = n$ et la fréquence observée pour chaque intervalle est $f_i = \frac{n_i}{n}$

Pour illustrer notre propos à partir de notre exemple, nous avons une variable dont l'unité est le kg. Essayons alors de concevoir ce qui intervient dans la construction de l'histogramme et comment nous pouvons interpréter.

unités des variables étudiées	*unités correspondantes sur le graphique*
variable “masse” en kg	unité de l’axe des abscisses : 1kg -------> 10 cm = 1 dm 0,1 kg = 1 hg ------> 1 cm
fréquence	unité d’aire 1%-------> 1 cm^2
variable “densité de fréquence”	unité de l’axe des ordonnées : 1%/kg --------> 0,1 cm 1%/hg ---------> 1 cm

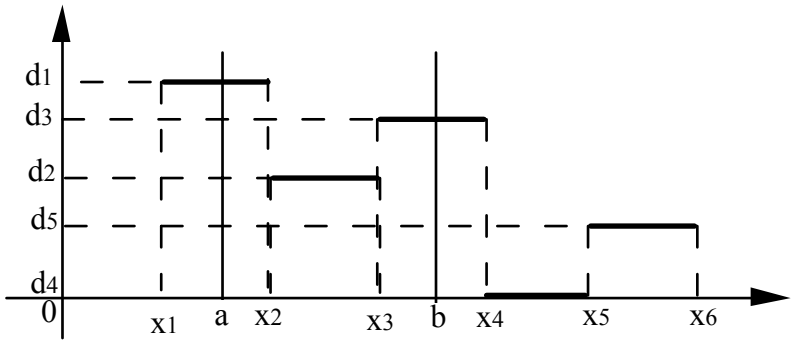
Le calcul de l’aire du rectangle ABCD revient à écrire la relation

$$(x_{i+1} - x_i)\, d_i = f_i = \frac{n_i}{n} \quad ,$$

de laquelle nous déduisons $d_i = \dfrac{f_i}{(x_{i+1} - x_i)} = \dfrac{n_i}{n(x_{i+1} - x_i)}$. Nous rappelons que le choix du nombre d’intervalles est arbitraire. Cependant il peut être fait de telle sorte que les amplitudes soient égales, c’est à dire $(x_{i+1} - x_i) = c$ pour i = 1 à k et même c = 1.

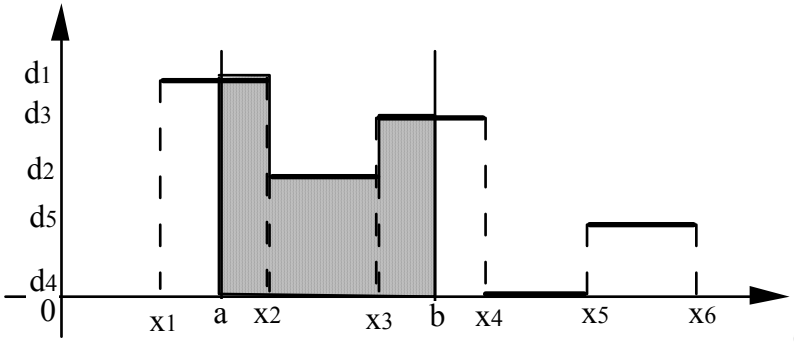
Dans notre exemple, nous pourrions convertir les mesures en hectogrammes et choisir les intervalles d’amplitude 1 hg. Ceci a pour conséquence que la valeur de d_i est égale à la valeur de f_i . C’est ici que naît l’ambiguïté, les deux variables ont même valeur mais ne se réduisent pas l’une à l’autre. Par boutade, ce n’est pas parce que dans une famille de quatre enfants, il y en a un qui est âgé de 4 ans que l’on est amené à confondre l’âge et l’effectif. Ou encore en roulant pendant une heure à vitesse constante 60 km/h, on parcourt 60 km, cela ne réduit pas pour autant la notion de vitesse à celle de distance. Or c’est ce genre de confusion qui est pratiquée dans de nombreux manuels scolaires ou divers ouvrages. Un des buts de cet article est de dénoncer cette confusion par souci de rigueur.

La fonction densité, c’est à dire ce que représente l’histogramme, que nous obtenons par observation statistique est alors une fonction constante par morceaux du type suivant:



(figure 3)

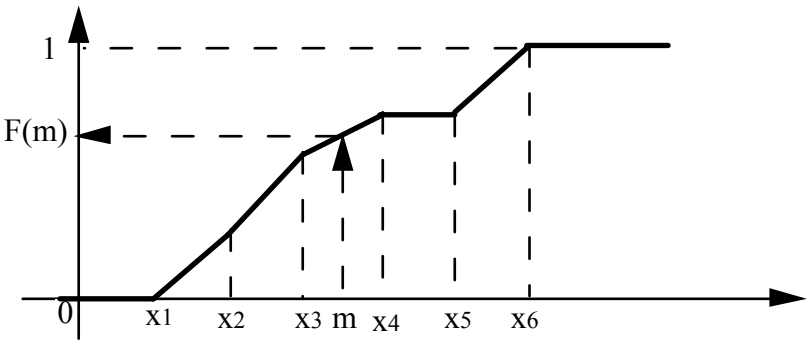
L'aire de la surface sous cette courbe (fig. 3) correspond à la fréquence. On peut alors estimer la fréquence possible des billes dont la masse est comprise entre les deux valeurs a et b en mesurant l'aire de la surface grisée (fig. 4).



(figure 4)

Il s'agit de calculer l'aire de trois rectangles. On pourrait aussi estimer la fréquence possible des billes dont la masse est inférieure à une valeur fixée quelconque. On obtiendrait la fonction cumulative croissante (fonction de répartition) qui n'est autre qu'une primitive de la fonction f.

La courbe est celle d'une fonction non décroissante affine par intervalle du type :



(figure 5)

$$F(m) = f_1 + f_2 + d_3\,(m-x_3) = P\,(X<m)$$

Pour quoi et comment peut-on lisser l'histogramme ?

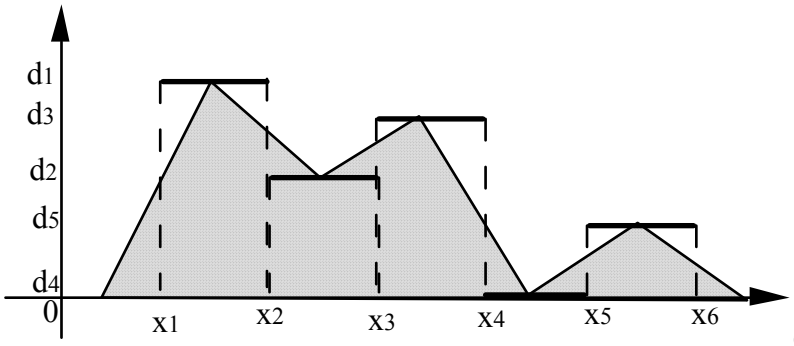
La procédure utilisée précédemment s'inscrit dans l'idée de chercher à estimer la loi de fréquence qui régit la variable "masse" en estimant la densité[5] à partir d'informations partielles issues d'un sondage. Dans le vocabulaire usuelle de la statistique apparaissent les termes de *polygone des fréquences*, du *polygone statistique* ou d'*ogive de Galton*. Les procédures

[5] on pourra consulter l'ouvrage *Histogrammes et estimation de la densité,* Delecroix, M., Que sais-je ? 1983

suggérées n'évoquent guère cette idée d'approximation et d'estimation. Il s'agit ici de remplacer l'histogramme représenté par une courbe constante par intervalle, par un autre histogramme représenté par une autre courbe enfermant une surface d'aire égale à la précédente (qui représente l'effectif total ou la fréquence 1) en supprimant les discontinuités (les sauts entre les segments horizontaux). Il y a évidemment une infinité de solutions à ce problème. Quand les intervalles sont de même amplitude la courbe (fig. 6) passant par les points successifs de coordonnées :

$A_0 (x_1 - \frac{x_2 - x_1}{2} ; 0)$ puis $A_i (\frac{x_{i+1} + x_i}{2} ; d_i)$ pour i = 1 à k-1 et enfin

$A_k (x_{k+1} + \frac{x_{k+1} - x_k}{2} ; 0)$



(figure 6)

A ce stade la nouvelle fonction densité dont l'histogramme est une représentation graphique est une fonction affine par intervalles qui est continue sur l'ensemble des nombres réels. On pourrait continuer à lisser. Mais on peut aussi se contenter de cette représentation pour la comparer à celle des variables connues servant de modèles telles que la variable de Laplace-Gauss pour ne citer que la plus utilisée. Le ou les maxima de la densité déterminent les valeurs modales de la variable étudiée.

Ainsi quel intérêt avons-nous à tracer un histogramme ?

Les quelques propriétés évoquées peuvent en quelque sorte servir à comprendre que l'intérêt de tracer l'histogramme ne se limite pas à l'obtention d'un graphique figuratif à des fins esthétiques et décoratives. Elles permettent aussi de s'opposer à un usage abusif de l'histogramme fondé en statistique descriptive sur une sorte de coutume poussant à tracer des graphiques sans toujours bien réfléchir à leur pertinence (ce qui est encore plus facile de nos jours avec l'usage de l'informatique) peut-être en pensant qu'il y a là une marque de scientificité. En respectant les propriétés qui président à sa définition, l'histogramme constitue un outil graphique permettant de présenter des données statistiques quantitatives issues d'une

variable continue ou même issues d'une variable discrète si ces données sont en très grande quantité. Pour illustrer ce dernier cas de figure, nous pourrions prendre l'exemple de la variable « nombre de *pile* obtenu en jetant 10000 fois une pièce de 5 francs ». Essayez de représenter sur l'espace usuel d'une feuille 21x 29,7, le diagramme en bâtons de la distribution des fréquences des 10001 résultats possibles. C'est d'ailleurs ce point de vue qu'avait adopté Laplace dans son traité de *Théorie analytique des probabilités* (1814). A côté de l'intérêt lié au domaine de la statistique, nous pourrions y trouver un intérêt pédagogique en explicitant les notions et méthodes mathématiques que la notion d'histogramme requiert et dont elle constitue une sorte d'exemple d'application. Nous reviendrons plus loin sur cette perspective.

La notion d'histogramme au travers des manuels scolaires ou d'ouvrages non spécialisés en statistique.

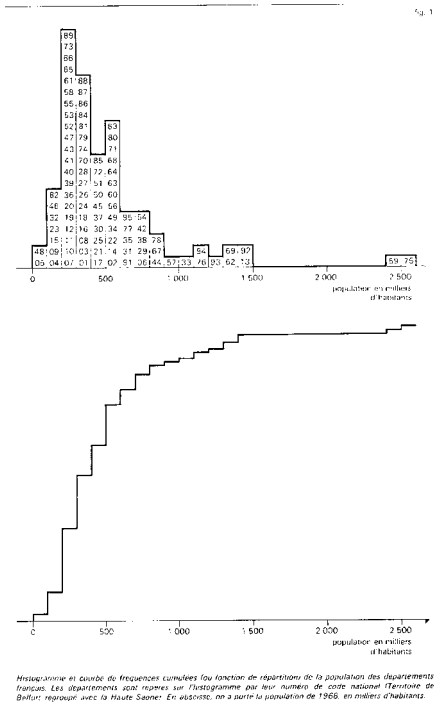
Dans *Le Grand Robert* [6] l'auteur donne la définition suivante « Graphique représentant la densité d'effectif en fonction des valeurs d'un caractère et formé par une série de rectangles dont la base constitue un intervalle de variation de ces valeurs et la surface, l'effectif correspondant » qui nous satisfait dans la mesure où elle introduit l'idée de *densité d'effectif*. En revanche *Le Petit Larousse illustré -1989* le définit comme « représentation graphique des classes d'une variable statistique associant à chaque classe un rectangle proportionnel par sa longueur à l'amplitude et par sa hauteur à l'effectif de cette classe » L'auteur l'illustre avec une figure qui ne correspond pas à un histogramme puisqu'il s'agit une variable chronologique des hauteurs en mm des précipitations par mois sur une année!

Comment une telle définition peut-elle conduire un apprenant vers la notion d'histogramme ? S'agit-il d'une contrainte ou d'une dérive de la vulgarisation scientifique ? Cette déformation résulte-t-elle du processus de transposition didactique, d'une négligence ou de l'ignorance ?

L'encyclopediæ Universalis rapporte à l'article Statistique[7] les deux figures (fig. 7) ci-contre

[6] article *Histogramme*, p 202, Vol 5 année 1986

[7] article STATISTIQUE, vol 15 -année 1980 p 328 (a)

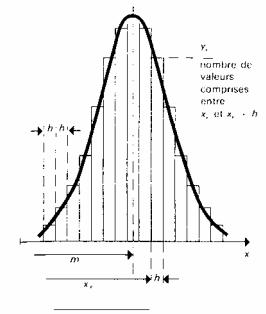


Histogramme et courbe de fréquences cumulées (ou fonction de répartition) de la population des départements français. Les départements sont repérés sur l'histogramme par leur numéro de code national (Territoire de Belfort regroupé avec la Haute-Saone). En abscisse, on a porté la population de 1968, en milliers d'habitants.

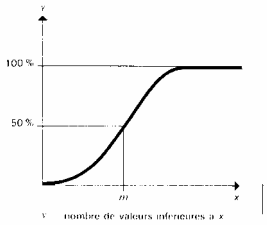
dans un encadré contenant selon la légende un histogramme et une courbe cumulée croissante (fonction de répartition) de la « population des départements français. Les départements sont repérés sur l'histogramme par leur numéro (...). En abscisse, on a porté la population de 1968, en milliers d'habitants ». Nous constatons que ces représentations graphiques ne se rapportent pas à des courbes de fonctions puisque les lignes de rappel sont intégrées à la courbe. Sous la rubrique *statistique descriptive*, on peut aussi lire « la représentation graphique d'un échantillon se fait à l'aide de l'histogramme, ou du polygone des fréquences. On utilise aussi le diagramme des fréquences cumulées ou fonction de répartition ».

A l'article Mesure[8], nous trouvons trois encadrés contenant respectivement un histogramme d'une série d'observations réalisé à partir de rectangles d'amplitude h (fig. 8), une courbe des fréquences cumulées (fig. 9) et la courbe de Laplace-Gauss (fig. 10)..

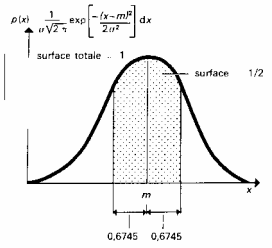
[8] article Mesure -Méthodologie de la mesure , vol 10 pp 853-854



Histogramme d'une série d'observations.



Courbe des fréquences cumulées.



Courbe de Laplace-Gauss.

La variable Y portée en ordonnée donne «y_k = nombre de valeurs comprises entre x_k et x_{k+1}» Sur l'histogramme (fig. 1) est posée la courbe de Laplace-Gauss suggérant un lissage. Un texte précise la procédure suivie : *« Dans la pratique, on groupe les résultats par valeurs croissantes. On calcule la moyenne m. On construit une courbe appelée histogramme en divisant la gamme des valeurs en tranches de largeur h et en rapportant, dans chaque tranche limitée par les valeurs x_k et x_{k+1}, le nombre de valeurs expérimentales correspondantes appelé aussi fréquences ».* Outre qu'une ambiguïté demeure quant au sens du mot fréquence : effectif ou proportion ?, le renvoi à l'illustration (fig.10) rappelle qu'il s'agit de la *densité de fréquence* et non de la *fréquence*. A moins que $p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-m)^2}{2\sigma^2}) = f(x)$ ne soit considérée comme l'expression algébrique donnant la fréquence (qui requiert la notation du type suivant : $dp(x) = f(x)dx$) et non la densité de fréquence du modèle gaussien. Dans la dernière édition dont nous avons extrait ces trois graphiques les deux figures 8 et 9 sont identiques à la précédente édition et sur la figure 10 la légende de l'ordonnée contient le symbole dx.

Enfin sur la courbe des fréquences cumulées (fig. 8) apparaît en ordonnée la variable Y correspondant à des % alors que la légende stipule « Y = nombre de valeurs inférieures à x ». Certes l'expert réalise immédiatement les rectifications. Et l'article met en évidence un des usages possibles de l'histogramme : un outil d'aide à la modélisation par ajustement d'une loi

de fréquence d'une variable à une loi théorique (ici celle de la variable de Laplace-Gauss). Mais il semble que le manque de rigueur est susceptible de maintenir une grande confusion chez le novice-apprenant soucieux d'une précision plus forte de la notion d'histogramme.

Nous avons aussi consulté de très nombreux manuels scolaires des classes de 4ème de collège

Terminale D - Géométrie & Statistique - Aleph 0 - Hachette 1972

programme du 14 mai 1971

A la page 194 la définition est fournie en ces termes :

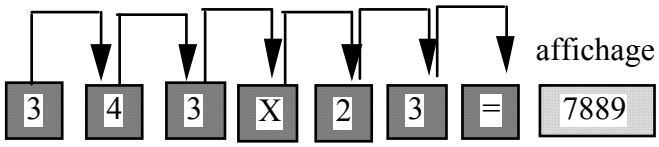
« On appelle histogramme, la représentation graphique d'une statistique d'un caractère dont les valeurs possibles ont été regroupées en classes d'égale amplitude, cette représentation étant obtenue de la manière suivante : sur chaque classe représentée par un segment de l'axe des abscisses, on construit un rectangle dont le côté parallèle à l'axe des ordonnées a une longueur proportionnelle à la classe envisagée ». Puis il est ajouté que « dans le cas où les classes ne sont pas d'amplitude constante, on peut néanmoins établir un histogramme en construisant des rectangles dont l'aire est proportionnelle à l'effectif de la classe envisagée». Pour le moins, l'expression " a une longueur proportionnelle à la classe envisagée" n'est pas dénuée d'ambiguïté. Mais nous pouvons penser que la confusion provient de ce que les auteurs s'astreignent à ne caractériser l'histogramme que par un algorithme de construction.

```
  343
x  23
------
 1029
 686*
------
 7889
```

C'est un peu comme si on définissait le concept de multiplication de deux nombres par la seule description d'un algorithme de calcul du type ci-contre ou encore par le mode d'emploi d'un boulier ou d'une calculatrice.

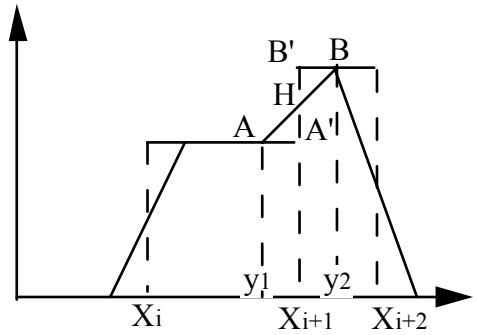
Cela pourrait donner :

On appelle « multiplication de deux nombres le résultat obtenu en appuyant successivement sur les touches :



La notion de "polygone statistique" est abordée en tant que « ligne polygonale obtenue en joignant les extrémités des bâtons du diagramme représentatif de la série statistique correspondante». Est ajouté à ce propos : « Dans le cas où les classes sont d'amplitude inégale, la construction du polygone statistique est plus délicate. Il faut se ramener au cas où les classes sont de même amplitude, de façon que, dans les intervalles séparant les

milieux de chacune de ces classes, l'aire limitée par la portion du polygone statistique concernée soit égale à l'aire du rectangle correspondant dans l'histogramme » . Cette remarque est judicieuse dans sa phase de mise en garde mais plutôt ambiguë dans l'expression de la contrainte.



(*figure 11)*

En observant de plus près, il s'agit de réaliser une construction géométrique (fig. 11) telle que l'aire du trapèze y1ABy2 soit égale à la somme des aires des deux rectangles $y_1AA'X_{i+1}$ et $X_{i+1}BB'y_2$. Le découpage des intervalles choisi donne les points y1 et y2 de telle sorte que les triangles rectangles AA'H et BB'H soient isométriques. Ici on trouve une situation d'application de la géométrie et l'on peut s'étonner qu'elle ne soit pas exploitée.

Nous constatons qu'un nouvel obstacle peut encore être introduit par le fait que le polygone statistique est défini par une procédure de construction non par une problématique comme celle du lissage par exemple. Nous ne retrouvons pas à propos de la statistique, l'exigence que les auteurs se sont assignés à l'égard de la géométrie. En effet nous pouvons lire en préface :

« *Nous pensons que la confusion entre espace vectoriel et espace ponctuel, admissible et même fructueuse à un niveau plus élevé, serait catastrophique ici ; elle conduirait à un retour en arrière fâcheux à un moment où l'on s'efforce à juste titre de toujours préciser exactement la nature des objets mathématiques que l'on est amené à manipuler* ».

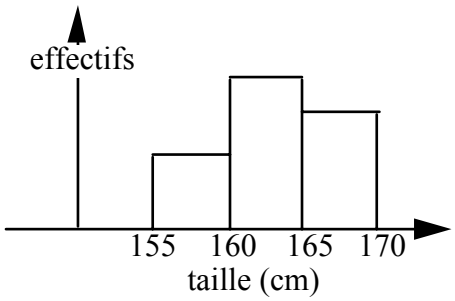
Terminale B Mathématiques -collection Durrande - 1980

programme 1971

A la rubrique *variable aléatoire numérique,* les auteurs introduisent la notion de *densité* par l'expression suivante : « P(X=x) est aussi appelée densité discrète de X en x »

Terminale B Mathématiques , C.Gautier, JC Martin, Hachette -1983

Dans le chapitre STATISTIQUE, la notion d'histogramme est ainsi abordée : « Dans le cas où l'on a opéré un regroupement en classes de valeurs de la variable, on utilise un histogramme». Un exemple est alors fourni au lecteur *les tailles en cm de trente élèves d'une classe de 1ère*. La représentation graphique fournie en guise d'histogramme n'est autre qu'un diagramme en bâtons puisque l'axe des ordonnées comporte l'étiquette *effectifs.*



(figure 12)

C'est ce genre de confusion que nous avons rencontrée dans de nombreux ouvrages. Nous reviendrons plus loin sur les conséquences de cette façon de faire.

Première CDE mathématiques - Tome 1 Monge et coll - Belin 1970

programme du 23 avril 1970

A la rubrique Histogramme on peut lire « on subdivise l'échelle des abscisses en intervalles consécutifs $[X_i; X_{i+1}[$ qui correspondent ... puis l'on construit sur chaque intervalle un rectangle dont l'aire est proportionnelle à la fréquence de la classe correspondante. Ces rectangles obtenus constituent l'histogramme de la distribution.» Puis est pointé le cas particulier : « lorsque toutes les classes ont la même étendue, on construit chaque rectangle en prenant une hauteur égale à la fréquence». La présentation de l'histogramme satisfait aux exigences de la notion mais le cas particulier génère à nouveau la confusion qui se trouve d'ailleurs confirmée par un dessin.

Première A mathématiques - H. Pochard -Gauthier-Vilars 1971

programme du 23 avril 1970

Quatre pages sont consacrées à l'histogramme. Après une étude préalable sur les effets d'un changement de repère cartésien orthogonal sur le calcul de l'aire d'un rectangle et un exemple d'*histogramme des effectifs*,

l'auteur propose la définition suivante : « Les histogrammes sont des diagrammes en surface où l'on représente les effectifs y_i (resp : les fréquences f_i) par des rectangles dont les aires sont proportionnelles aux y_i (resp : les fréquences f_i)»

Première F mathématiques - Col DIMATHEME - Didier- 1988

programme de 1985

Par définition « l'histogramme des effectifs est formé de rectangles ayant des aires proportionnelles aux effectifs des classes ». Cependant la confusion réapparaît avec le cas des intervalles d'amplitude égale où l'axe des ordonnées devient l'axe des effectifs.

Seconde mathématiques algèbre -Audirac, Bonvallet- MAGNARD - 1988

programme 1981

Cet ouvrage ne comporte pas de définition explicite mais à la rubrique *variables statistiques continues* apparaît une procédure de construction et un graphique conformes à notre définition. L'indication de l'unité d'aire en légende en témoigne.

Seconde - mathématiques - Glaymann, Malaval -CEDIC 1981

programme 1981

Dans un chapitre intitulé "analyse des données", l'histogramme est présenté conformément à notre définition. Il est abordé par une série statistique dont les classes sont d'inégale amplitude puis les auteurs précisent : « Dans le cas particulier où les classes sont d'égale amplitude, le calcul des hauteurs des rectangles est simplifié : il suffit de prendre pour ordonnée des valeurs proportionnelles aux effectifs». Et ils ne proposent pas de mettre en ordonnées les effectifs !

Seconde - mathématiques IREM Strasbourg -ISTRA- 1981

programme 1981

L'histogramme est abordé dans le thème Graphiques divers. Il est conforme à notre définition comme le confirme la légende donnant l'information 4% pour dire 1 cm^2 correspond à 4%. L'exemple choisi porte sur une variable discrète "nombre de lettres des mots d'un texte". Ce qui

n'est peut-être pas très pertinent sauf à réaliser une extension d'une variable discrète à une variable continue. Mais cette perspective n'est pas explicitée.

Seconde - mathématiques - coll LOUQUET -Armand Colin- 1986

programme de 1986

Dans le paragraphe consacré au cas d'*un caractère continu - les diagrammes à bandes*

« Dans le cas où l'on dispose de classes statistiques, les intervalles des classes (qui ne sont pas forcément tous de même amplitude) sont reportés sur un axe. Ils servent de bases à des rectangles dont les aires sont proportionnelles aux effectifs.» Un dessin est alors proposé sans l'axe des ordonnées. Ce qui nous semble être un moyen didactique pertinent pour éviter la dérive que nous dénonçons qui conduit à l'interprétation de l'axe des ordonnées comme axe des effectifs. « Le diagramme obtenu s'appelle un histogramme». A cela les auteurs ajoutent une mise en garde précieuse : « Attention, dans le cas où les classes sont d'amplitudes inégales, à bien calculer les hauteurs pour que ce soit les aires qui soient proportionnelles aux effectifs»

Seconde - mathématiques - coll Perspectives -Hachette- 1990

programme de 1990

L'histogramme est présenté par sa construction à partir de l'exemple de la variable *"prix d'une robe exprimé en francs"*. Le graphique est conforme à notre définition et comporte la légende 1cm2 représente 20 robes . La courbe des fréquences cumulées croissantes jouxte l'histogramme. Elle est cependant incomplète dans la mesure où elle est limitée aux point extrêmes observés.

Troisième - mathématiques - IREM de Strasbourg -ISTRA- 1993

Le chapitre 14 intitulé STATISTIQUES (au pluriel) comporte la rubrique *savoir présenter des données statistiques et les exploiter*. Cette rubrique explicite un objectif de savoir faire. Aucune définition explicite de l'histogramme n'est fournie mais un problème[9] est proposé. A partir de l'étude de la variable *"durée de vie des ampoules mesurée en heures"*, on demande de «représenter le tableau par un histogramme (un diagramme en

[9] extrait de Brevet des Collèges de Lille -1990

barres)» en donnant la contrainte « on prendra 1 cm pour 100 heures et 1 cm pour 100 ampoules».

Pourquoi confondre ces deux notions ?

Le terme diagramme en barres ne suffit-il pas ?

Où peut donc bien apparaître l'idée de la prise en compte de la surface dans une telle approche ?

Reportons à un autre exercice[10]. Il s'agit d'une série statistique portant sur la variable durée du trajet entre le collège et le domicile mesurée en minutes. Le résultat est rapporté graphiquement par un histogramme (le mot est employé). Cependant si l'axe des abscisses est bien l'axe des durées avec des classes d'amplitudes 5 minutes (1 cm), l'axe des ordonnées est désigné comme l'axe des effectifs. Évidemment malgré ce défaut il est possible de répondre aux deux questions posées :

1) Combien y a-t-il d'élèves dans ce collège ?

2) Combien d'élèves mettent moins d'un quart d'heure ? Quel pourcentage cela représente-t-il ?

En effet celles-ci n'impliquent que des classes entières. Il suffit donc d'utiliser les effectifs de chaque classe qui, pour des raisons purement techniques, coïncident avec les valeurs représentant les hauteurs des rectangles.

Mais nous aimerions savoir comment les auteurs répondraient aux questions suivantes:

1) Quel pourcentage d'élèves mettent plus de 7 minutes et moins de 2 minutes ?

2) Quel pourcentage d'élèves mettent moins de 23 minutes ?

3) Quel pourcentage d'élèves mettent exactement 23 minutes ?

Avec le modèle proposé, nous serions curieux de connaître la réponse des auteurs du problème en prenant appui sur ce qu'ils dénomment *histogramme*.

Quatrième - mathématiques - IREM de Strasbourg -ISTRA- 1993

Le chapitre 15 intitulé STATISTIQUES (au pluriel) comporte l'activité sur le "Tracé d'un diagramme en barres". Les auteurs justifient ce choix par « la plupart du temps, une enquête est accompagnée d'un dessin appelé diagramme» Ce texte nous paraît mettre davantage l'accent sur la

[10] n° H page 217

fonction illustratrice et esthétique du diagramme que sur sa fonction d'expression d'une notion dans un registre sémiotique graphique dans le but d'en faciliter le traitement. Après quoi les auteurs livrent l'algorithme de construction du diagramme. « ... notre enquêteur a commencé le diagramme ci-dessous constitué d'un rectangle dont la *base* est la largeur de la tranche de prix concernée et la *hauteur*, le nombre d'élèves correspondant (on dit l'effectif)» Certes les auteurs emploient le terme de diagramme en barres et non celui d'histogramme alors que la description suggère celle d'un histogramme. Ceci se confirme car ils nomment diagramme en bâtons la représentation graphique de la distribution des effectifs d'une variable quantitative discrète. En voulant sans doute éviter le terme, les auteurs introduisent une ambiguïté dont nous aimerions apprécier les conséquences. Faut-il voir dans le diagramme en barres une représentation qui n'est déjà plus un diagramme en bâtons sans être encore un histogramme? Est-ce un effet de transposition didactique aménageant une sorte d'espace transitionnel ?

Continuons l'exploration de l'exemple. L'aire du rectangle qu'ils proposent en modèle mesure 4,5 cm^2 et représente 9 individus si on lit l'ordonnée, cependant au sens de l'histogramme sur l'intervalle [2 ; 2,5[la densité d'effectif vaut constamment 9 individus par 0,5 F soit encore 18 individus/Franc. Il est alors demandé de reproduire et compléter ce diagramme appelé diagramme en barres puis de fournir le nombre d'élèves et le pourcentage parmi les élèves interrogés prêts à payer au moins 3F. Force est de constater que la seconde question n'exige en aucune façon l'usage du diagramme qui demeure dans sa fonction illustratrice.

Pourquoi et dans quel but n'a-t-on pas transformé cette variable en une variable qualitative ordinale à 5 modalités pour laquelle le diagramme en barres serait devenu avec pertinence un diagramme en bâtons ? Ce qui pourrait alors être l'occasion d'un lecture graphique de la modalité dominante (mode). Nous ne percevons pas où l'activité de lecture d'une représentation graphique peut se déployer.

De toute évidence les auteurs ont souhaité introduire une représentation géométrique prenant en compte la surface. Mais pourquoi en évitent-ils l'usage ?

Une question pertinente pourrait être : A combien peut-on estimer le nombre d'élèves qui sont prêts à payer entre 2,25 F et 3,75 F ? Celle-ci suppose deux conditions. D'une part que l'on introduise l'idée de la perte

d'information initiale, sinon il suffit de retourner à la série statistique initiale et de dénombrer l'effectif de l'intervalle [2,25; 3,75[. D'autre part qu'une information inconnue est remplacée par une estimation. Cette opération d'estimation peut être mise en relation avec des pratiques usuelles de chacun face des situations d'ignorance. Il nous semble que nous entrons ici de plain pied dans le champ de la statistique. La question ci-dessus introduit l'intérêt d'une représentation en surface et en impose une lecture adéquate. Quoiqu'il en soit, nous ne voyons pas ce qui s'oppose à l'introduction d'une représentation par surface proportionnelle à l'effectif ou à la fréquence. Les opérations de construction se trouveraient ainsi modifiées : « ... le diagramme... est constitué d'un rectangle dont la base correspond à la largeur de la tranche de prix concernée et l'aire correspond au nombre d'élèves...» Ces consignes conduisent à des tâches plus riches portant sur des objets plus pertinents dans le domaine de la statistique mathématique.

Réflexions sur le traitement de la notion d'histogramme dans l'enseignement de la statistique comme branche des mathématiques.

Certes il est possible de trouver des manuels ou des traités de mathématiques et de statistique qui se conforment à une définition cohérente de la notion d'histogramme. Mais force est de constater que les remarques que nous avons faites ne se limitent pas à l'échantillon des ouvrages que nous avons pris ici en exemple. Or imaginons un étudiant ou un enseignant souhaitant au moins s'informer sur la notion, il sera très probablement confronté à la situation que nous avons dénoncée. Ainsi comment ce novice pourra-t-il se faire une idée correcte de cette notion ?

Mais alors sur quoi se fonde la présentation de l'histogramme choisie par les auteurs en particulier celle qui interprète l'axe des ordonnées comme axe des effectifs ou des fréquences ?

Nous pouvons penser que cela tient à l'ignorance des auteurs ou plutôt à leur désintérêt pour la notion. Celle-ci se trouverait présentée dans les ouvrages uniquement parce qu'elle doit y figurer au nom des programmes scolaires. A moins que cela tienne à une phénomène de transposition didactique, c'est à dire compte tenu du fait que la notion implique des concepts de mathématiques avancées, il est réduit à ses caractéristiques élémentaires explicitées au travers d'un algorithme de construction géométrique. Qui plus est, le choix d'intervalles de même amplitude constitue un modèle adapté à de nombreuses situations de mesure

qui se trouve renforcé par la facilité de la construction. Ainsi le pédagogue estime lever un obstacle à l'apprentissage rapide de la notion d'histogramme en évitant l'affrontement à une représentation graphique surfacique qui doit être lue simultanément dans ses deux dimensions. Le problème est qu'alors la propriété particulière d'égalité numérique de la hauteur et de la fréquence (ou resp de l'effectif) est rapidement convertie en identité conceptuelle des deux objets: "La hauteur représente l'effectif, la fréquence ou la probabilité". Ceci se produit à l'insu de l'apprenant.

Nous pensons que cette approche didactique est génératrice d'obstacles didactiques.[11] Elle contribue à la construction d'une représentation mentale chez le sujet apprenant qui fait obstacle à la compréhension de l'histogramme quand il est amené à en faire usage dans un sens autre que celui d'une illustration fonctionnant comme un diagramme en bâtons. Nous l'avons maintes fois constaté dans le cadre de l'enseignement de la statistique à l'université[12]. Nos étudiants expriment leur étonnement ou déclarent leur incompréhension complète quand nous leur expliquons que le *hauteur* (c'est à dire l'ordonnée d'un point) dans l'histogramme ne correspond pas à la *fréquence* mais à une *densité de fréquence*. Le coût pour dépasser cet obstacle nous paraît plus élevé dans ce contexte que dans celui de l'apprentissage initial. L'approche de la construction de *l'histogramme* à partir de classes d'amplitudes inégales suivi de la construction d'un *lissage* du type *polygone statistique*, nous paraît être souhaitable dès la classe de Quatrième. Il y a une occasion d'activités mathématiques mettant en jeu divers cadres de référence : cadre géométrique, cadre numérique, cadre algébrique, cadre statistique, cadre de l'analyse fonctionnelle, mais aussi divers registres sémiotiques de représentation : registre graphique, registre discursif, registre algébrique.

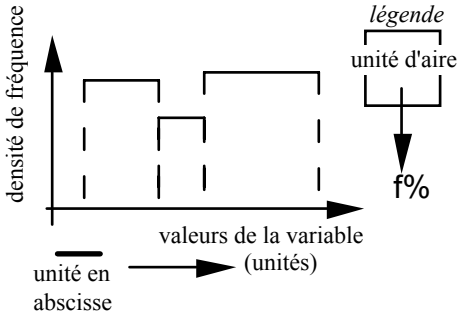
Par exemple pour ne pas induire l'identité *histogramme = rectangle*, nous suggérons des tracés du type suivant :

---

[11] Au sens de Bachelard, G., qui parle d'obstacle pédagogique en 1938 dans *la formation de l'esprit scientifique* (p 18 VRIN-1989)

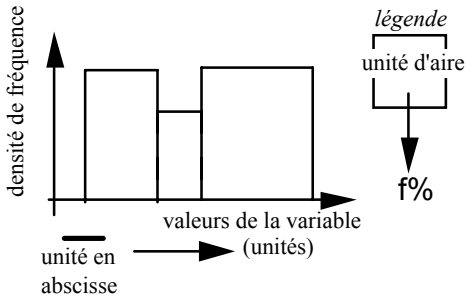
Au sens de Brousseau, G., qui donne un exemple à partir de l'enseignement des nombres décimaux fondé sur les mesures avec les unités et les sous-unités. Cette approche conduit des individus à concevoir les nombres décimaux comme des nombres constitués de deux parties autonomes : partie entière et partie décimale dont on retrouve les traces dans des erreurs du type $(0,2)^2 = 0,4$ mais non détectable pour $(0,5)^2 = 0,25$.

[12] en licence et en maîtrise de sciences de l'éducation



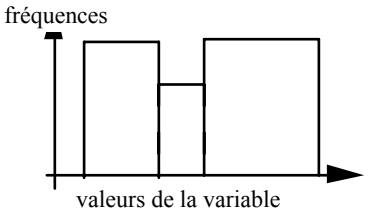
((figure 14)

au lieu de



(figure 15)

Et naturellement nous rejetons les tracés qui ne donnent pas les bonnes informations minimales. Ainsi nous bannissons cette représentation graphique (fig 16)



(figure 16)

A côté des manuels, il ne faudrait pas oublier les logiciels qui donnent un accès facile aux traitements en permettant des constructions automatiques de diagrammes nommés histogrammes mais qui ne sont, à ce que nous connaissons à ce jour, que des diagrammes en bâtons. C'est le cas d'un intéressant et pratique logiciel de traitements statistiques tel que CHADOC[13] dans lequel l'auteur désigne les diagrammes par histogrammes. C'est aussi le cas d'un tableur tel que Excel dont les possibilités sont immenses. Dans sa version actuelle comportant l'opération de traçage des courbes en coordonnées cartésiennes, il est possible de procéder à la réalisation de n'importe quel *histogramme* en considérant qu'il s'agit de tracer la courbe d'une fonction constante par intervalle puis celle d'une fonction affine par intervalle pour l'obtention de *polygone statistique*. Encore convient-il que les utilisateurs aient acquis quelques connaissances mathématiques qui fondent ces notions pour obtenir ce résultat avec un ordinateur, car ce dernier ne pense pas à la place du sujet, il ne fait que des traçages automatiques! Notons que dans ce cadre des traitements statistiques assistés par informatique, il y a de nombreuses autres réflexions et recherches à conduire en lien avec la multiplication des recours aux représentations graphiques les plus variées réalisées automatiquement sans aucun effort par l'utilisateur et même sans nécessité de compréhension. En

[13] IUT de Nice

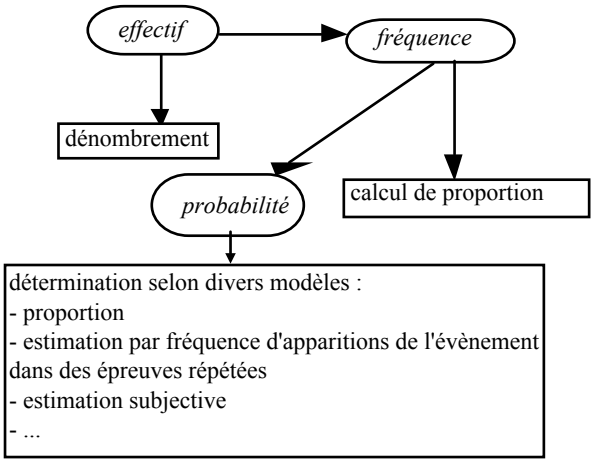
prenant les perspectives nomographique, épistémologique et heuristique dans le cadre de la statistique, à quoi servent en fait les représentations graphiques ?. Dans les cadres didactique et pédagogique, comment peut-on enseigner à construire, lire ou choisir une représentation graphique, que celle-ci soit à produire dans la classe ou qu'elle soit produite ailleurs et introduite en classe par l'intermédiaire des *media* ?

Il nous semble tenir là un exemple de notion mathématique introduite par l'enseignement en dehors de toute problématique pertinente. Face à cela, nous pouvons adopter deux conduites :

- ne pas l'enseigner

- l'enseigner sur la base de problématiques pertinentes accessibles à des élèves des différents niveaux scolaires.

Nous rappelons que notre critique à l'égard de l'enseignement de l'histogramme ne se fonde nullement sur un culte particulier auquel nous vouons cette notion. C'est tout le contraire. Nous souhaitons proposer cette notion dans un réseau notionnel pour éviter une dérive d'un enseignement de l'histogramme pour lui-même. L'histogramme est un objet mathématique qu'il faut éviter de réifier. Il importerait de préciser selon les circonstances s'il est traité comme une notion-outil ou comme une notion-objet. Examinons maintenant dans quel réseau notionnel nous pourrions l'immerger. Cette connaissance peut alors servir pour construire des séquences didactiques mettant en jeu diverses notions, méthodes et concepts mathématiques à des degrés variés d'abstraction. Elle peut aussi conforter l'idée qu'il ne s'agit pas d'une notion greffée arbitrairement.



Trois notions sont impliquées *effectif* , *fréquence* et *probabilité* qui renvoient aux opérations ci-contre. A celle de *fréquence* est souvent attachée celle de *pourcentage*..

Examinons maintenant quelques jeux de cadre possibles. La définition de l'histogramme peut être exprimée:

- dans un cade géométrique :

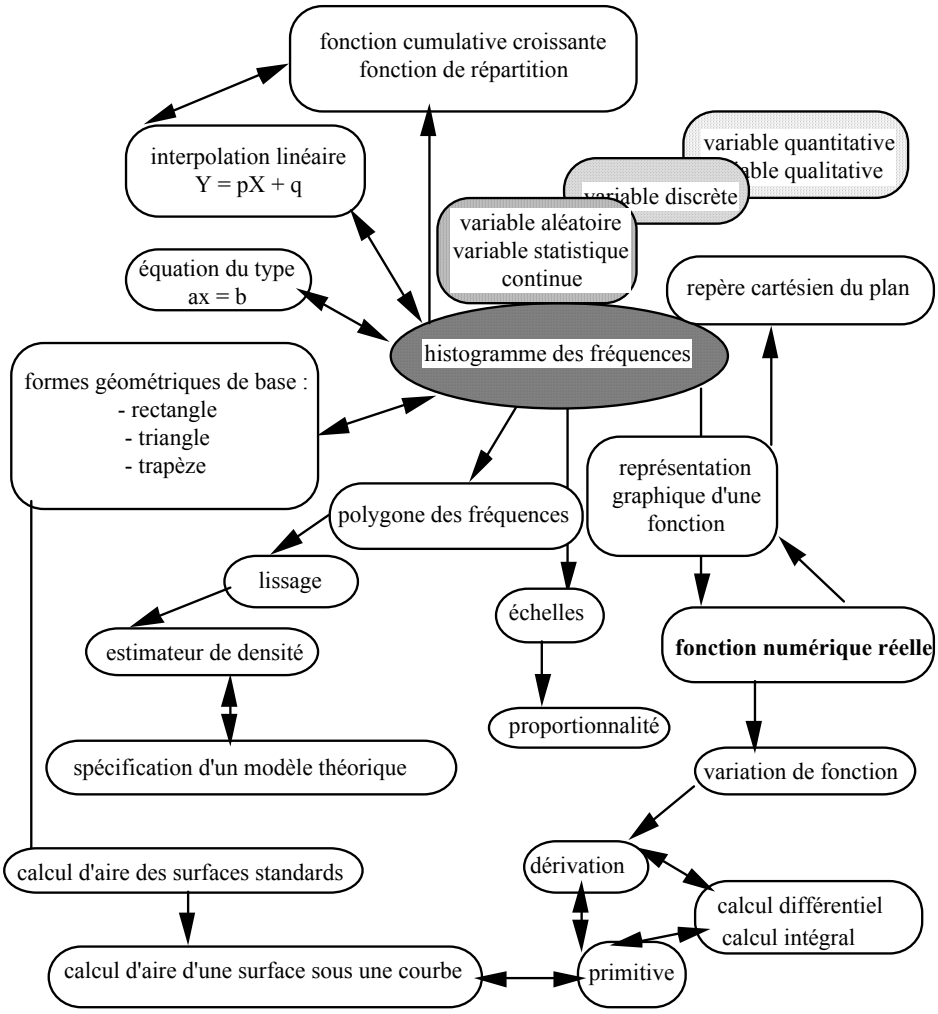
« Un histogramme est une surface dont l'aire totale est égale à 1...».

- dans un cadre géométrique analytique :

« Dans un repère cartésien orthogonal, un histogramme est une surface délimitée par une courbe représentative de la densité de fréquence d'une variable statistique continue dont l'aire totale est égale à 1...»

- dans un cadre algébrique :

« un histogramme est l'ensemble des couples (x ; y) tels que pour x nombre réel de l'ensemble définition Df, $0 \leq y \leq f(x)$, et où x est une valeur possible, résultat de la variable statistique et f(x) est la densité de fréquence de cette valeur x ...»



Lorsque f(x) est une expression algébrique connue, nous sommes ramené à un problème classique d'étude d'une fonction numérique de variable réelle prévue dans les programmes scolaires. Dans l'introduction habituelle de l'histogramme, il s'agit tout simplement de réaliser l'étude

d'une fonction constante par intervalle (fonction en escalier). Les intervalles étant caractérisés par la situation étudiée.

En ce qui concernent les registres sémiotiques, nous avons à utiliser du texte, des écritures symboliques, des tableaux et des représentations graphiques géométriques.

Pour quelles raisons et dans quels buts l'enseignement habituel de l'histogramme n'établit-il pas clairement ces liens et ne met-il pas les apprenants dans des situations propres à établir les correspondances et les diverses traductions qu'elles imposent ?

Quels avantages cet enseignement tire-t-il en isolant la notion d'histogramme du cadre important des fonctions ?

Ne peut-on pas dès la classe de quatrième introduire des situations qui évitent de retirer à l'histogramme sa finalité majeure qui est de représenter des variables par des surfaces dont les aires traduisent les valeurs ? Remarquons que le problème attaché au calcul de l'impôt sur le revenu est du même type.

Si nous imaginons que le modèle initial d'histogramme est celui présenté en classe de quatrième[14], comment quelques années plus tard un étudiant pourra-t-il comprendre que l'histogramme de la variable de Laplace-Gauss ("la courbe en cloche") se rattache à la même idée que celle développée à propos de la distribution des effectifs de la variable *"prix d'une barre glacée"* ?

Le lecteur aura compris qu'au-delà des critiques fondées sur les observations de nos étudiants en cours d'apprentissage de statistique et des outils d'aide à l'apprentissage que sont les manuels, les dictionnaires ou les logiciels, nous avons posé plus de questions que nous avons apporté de réponses. Toutefois la question formulée est un premier pas vers la résolution du problème qu'elle traduit. Notre objectif est avant tout d'attirer l'attention des enseignants de mathématiques. Mais la notion d'histogramme est aussi l'affaire de bien d'autres disciplines scolaires et il serait utile de connecter notre réflexion à d'autres, semblables, conduites en géographie, en physique ou en biologie par exemple. Nous demeurons certain que des situations didactiques pertinentes peuvent être organisées par des enseignants. C'est d'ailleurs à ce prix que nos conjectures seront validées ou réfutées et que notre propos ne sera pas dogmatisé.

---

[14] *op. cit.*